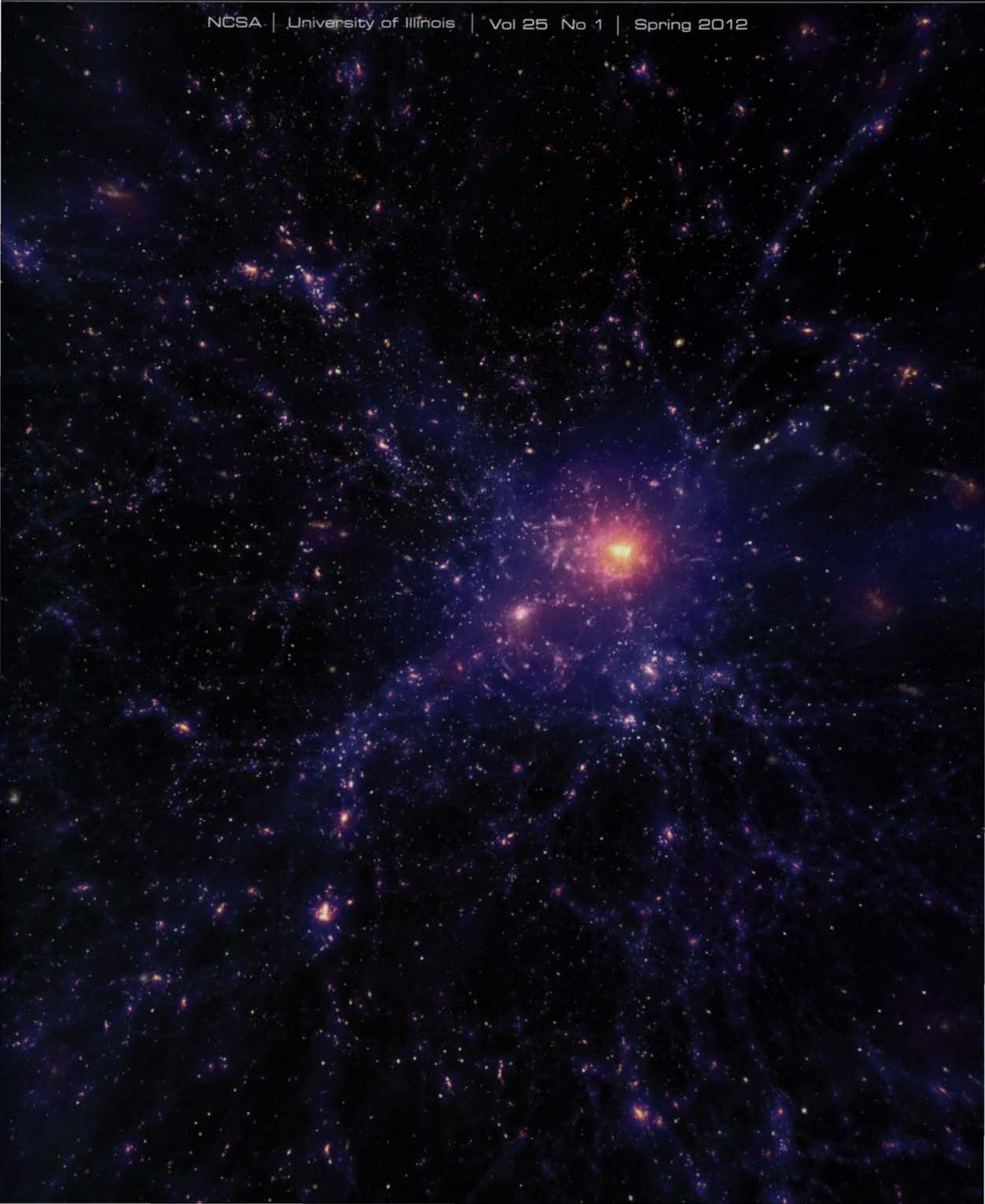
access

NCSA | University of Illinois | Vol 25 No 1 | Spring 2012



NCSA

access

Vol 25 No 1 Spring 2012

Managing Editor
Barbara Jewett
barbaraj@ncsa.illinois.edu

Art Director
Paula Popowski
popowski@ncsa.illinois.edu

Copy Editor
Trish Barker

On the Web

News from NCSA is always available on the Web. Sign up for our RSS feed or email service for the latest developments from NCSA and its partners.

www.ncsa.illinois.edu

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
1205 West Clark Street
Urbana, IL 61801
217-244-0072

Who we are

The University of Illinois at Urbana-Champaign's National Center for Supercomputing Applications (NCSA), one of the five original centers in the National Science Foundation's Supercomputer Centers Program, opened its doors in January 1986. Over the years NCSA has contributed significantly to the birth and growth of the worldwide cyberinfrastructure for science and engineering, operating some of the world's most powerful supercomputers and developing the software infrastructure needed to efficiently use them.

That tradition continues as the center, Illinois, Cray, and their partners in the Great Lakes Consortium for Petascale Computation develop what is expected to be the first computer dedicated to open scientific research capable of sustaining more than one petaflop, or one quadrillion calculations per second. Called Blue Waters, the system will be dedicated to massive simulations and data analysis projects that will improve our society, health, environment, and economic competitiveness. NCSA and the consortium will also work with research communities to create the new software technologies, scientific applications, and educational programs needed to take full advantage of this new system.

Blue Waters will benefit from NCSA's ongoing focus on cyberenvironments, cyber-resources, and innovative systems research. Cyberenvironments give research communities the means to fully exploit the extraordinary resources available on the internet (computing systems, data sources and stores, and tools). Cyber-resources ensure computing, data, and networking resources are available to solve the most demanding science and engineering problems and that the solutions are obtained in a timely manner. Innovative systems research involves testing and evaluating the performance of emerging computing systems for scientific and engineering applications.

NCSA also leads the Extreme Science and Engineering Discovery Environment (XSEDE), the most advanced, powerful, and robust collection of integrated advanced digital resources and services in the world. It is a single virtual system that scientists can use to interactively share computing resources, data, and expertise. Scientists and engineers around the world use these resources and services—supercomputers, collections of data, and new tools, for example—to make discoveries that benefit society. The five-year, $121 million project is supported by the National Science Foundation. XSEDE replaces and expands on the NSF TeraGrid project. More than 10,000 scientists used the TeraGrid to complete thousands of research projects, at no cost to the scientists.

The center also leaves its mark through the development of networking, visualization, storage, data management, data mining, and collaboration software. The prime example of this influence is NCSA Mosaic, which was the first graphical web browser widely available to the general public. NCSA visualizations, meanwhile, have been a part of productions by the likes of PBS's NOVA and the Discovery Channel. Through its Private Sector Program, top researchers explore the newest hardware and software, virtual prototyping, visualization, networking, and data mining to help U.S. industries maintain a competitive edge in the global economy.

Support for NCSA is provided by the National Science Foundation, the state of Illinois, industrial partners, and other federal agencies. For more information, see www.ncsa.illinois.edu.

Cover

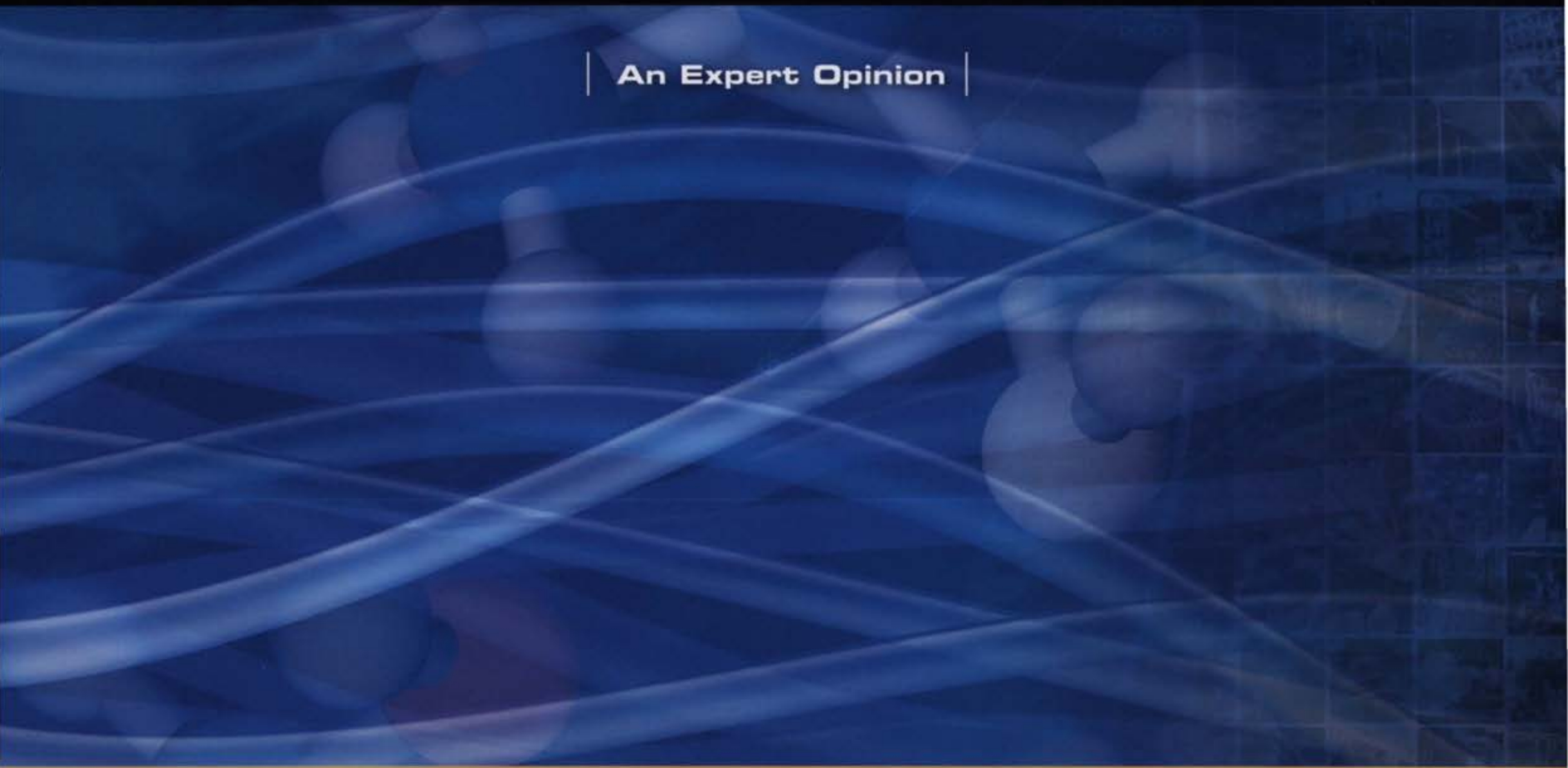
The front cover (and pages 28-29) portrays the evolution of filamentary structure in the early universe, using data from Princeton astrophysicists Renyue Cen and Jeremiah Ostriker. On the back cover is a dramatic collision of galaxies created based on simulations conducted by Brant Robertson at the University of Arizona. All images were created by the Advanced Visualization Laboratory at NCSA.

access

contents

02 **NCSA 2015**

Thom Dunning
Director, NCSA



NCSA 2015

EVERY FIVE YEARS NCSA UNDERTAKES THE DEVELOPMENT of a new strategic plan. The strategic planning process gives NCSA leadership and staff an opportunity to critically examine our activities during the past five years as well as look forward to where research, education and technology are going in the next five years. We began the planning process for NCSA 2015 in 2010 and are now in the process of implementing elements of the new plan. Below is a brief discussion of two major outcomes of NCSA 2015: the new Advanced Digital Services and Data-intensive Computing directorates.

Advanced Digital Services. In the next five years, NCSA expects to be operating computing and data systems for a large number of projects, including Blue Waters, XSEDE, the Dark Energy Survey, and the Large Synoptic Survey Telescope. NCSA will also be providing resources and services to researchers and educators in the K-16 education system, industry and government. The range of research areas to be served will be broad: from astronomers and biomedical researchers through chemists and environmental scientists to epidemiologists and ecologists and social scientists, historians and choreographers. These communities will require an equally broad range of support, from operation of a diverse set of computing and data resources and services, to help desk, application programming assistance, and advanced user support.

NCSA has been successful in meeting the needs of researchers and educators in the past. But, with many more research and education fields now needing high-performance computing and data resources and services, the scope, range, and intensity of future demands will greatly exceed those that have been experienced thus far. NCSA decided to meet this need for expanded resources and services by redesigning its computer and data service operations, organization, and management.

The new Advanced Digital Services (ADS) directorate will enable NCSA to provide the services and resources needed by a broad range of researchers and educators. ADS will provide NCSA with the flexibility needed to provide the technical skills for its projects, retain skilled staff as funding streams come and go over time, and encourage staff to collaborate and share knowledge with their counterparts working on other projects. We are now seeking an individual to lead this new effort.

Data-intensive Computing. One of the most exciting advances in science and engineering in the past decade is the fundamental shift in the use of data that was born digital. Fields as disparate as astronomy, biology and engineering, environmental and geosciences are being revolutionized by the use of digital technologies: digital cameras in astronomy, highly automated sequencers in biology, and sensor arrays in engineering, environmental and geosciences. Data analysis and data-driven discovery build on advanced information systems to collect, transport, store, manage, integrate and analyze these increasingly large amounts and diverse types of raw data and the resultant data products to make them accessible, searchable and usable to the wider community of researchers, which can include citizen scientists. The knowledge gained from data analysis and

data-driven discovery is already transforming our understanding of natural and societal phenomena, and the future is full of promise.

NCSA has always had significant efforts in data-intensive computing, but they were fragmented—a collection of individual projects, not an integrated program. As a result of the strategic planning process, we established the Data-intensive Computing directorate, mirroring our other major programs in extreme-scale computing and national cyberinfrastructure. We are now establishing the initial direction and structure of the new program.

There were many other conclusions from our development of NCSA 2015. We will also be implementing these recommendations over the rest of this year.

Thom Dunning
Director, NCSA

SERVING THE WORLD OF SCIENCE

Alan Blatecky, head of the National Science Foundation's Office of Cyberinfrastructure, recently chatted with Access' Barbara Jewett about the new visions at OCI and NSF.

Q. The Advanced Computing Infrastructure plan has five strategies. Is there one that you feel is more important or more urgent to address than the others?

A. We carefully selected those five. It is important to understand what we're looking at an ecosystem, if you will. One of the problems we've had in the past is people too focused on one thing or the other. If you look at what we're doing in CIF 21 [Cyberinfrastructure Framework for 21st Century Science, Engineering and Education], you really can't effectively focus on one strategy without doing things in other areas. For example, the first ACI strategy talks about foundational research in parallelism and concurrency and so forth. That is absolutely critical, but the point is you then need to connect that to the science being done, which is what the second strategy is about. The third strategy says that while you are working on the first two strategies, you also have to have infrastructure to work on, and if you have that then you need expertise and people. So which one is more important depends on which hat you're wearing at the moment.

Q. One of the criticisms that many in the HPC community often level at NSF is that NSF funds hardware, but not software development. CIF 21 has software included in its framework. And applications development is ACI strategy number two. Is this signaling a shift in NSF's thinking?

A. Yes, you are absolutely right on target. Software and data are becoming ever more important in computational science. Several years ago, data was often considered a by-product of science. Today data and software are absolutely essential to every science; in fact, it seems pretty clear that advances in computational capabilities and innovation require significant research efforts in software and algorithm development.

Q. As I read through the ACI strategies, it struck me that they seemed to be describing multi-site, multi-resource collaborations like the current XSEDE project. Am I correct in that? You're bringing everybody to the table?

A. Right. This is because some issues and problems can only be solved by involving multiple disciplines. EarthCube is a good example. This is a major new initiative and partnership OCI has with the Directorate for Geosciences. We brought together a wide range of communities to better understand the earth from its center to the sun. This meant that we had to engage a broad range of geoscientists and infrastructure developers. To kick this off, we conducted a four-day charrette with these communities. At the end of the charrette, 60 percent of those who participated said they needed data outside their own area and discipline to do the work they needed to do. [Editor's note: A charrette (pronounced shuh-ret) is a method—usually a series of meetings—to capture the vision, values, ideas, and talents of all interested parties with the goal of creating and supporting a feasible plan.] When we start looking at ways to address grand challenge problems such as this, it is evident that they require an enormous amount of capabilities and a wide range of expertise.

Q. Where do the NSF-funded centers fit into this plan?

A. If the centers are not already looking seriously at the ACI strategies, they ought to do so now. The NSF centers need to effectively participate in all these areas to address the needs of science. It's not just having a computational resource, it's about supporting the entire computational ecosytem. The challenge is how do you leverage these to support science and education.

Q. Tell me more about strategy four, the education and workforce programs. How do you envision this being implemented?

A. That, I'll be honest, is one of the toughest things we have to do. Because ACI is an integrated ecosystem, it requires people with expertise in many different areas. For example, we have already talked some about the importance of computational science and data-enabled science. We need that sort of expertise in order to do science, and that is much more difficult than saying "Oh, let me give you a course in how to write a new algorithm." It's much more complex.

So the issue is how to do we build a new education and workforce pipeline for computational and data-intensive science. We need to have it extend across the entire spectrum; from the faculty and the researcher, to students and post-docs, to the technician. How do we begin to expand and develop that much larger and complex education and workforce capability? This has to also include addressing how do I understand the problem and construct it in a way I can solve it.

Q. Many have said GPUs are the way of the future for scientific computing, and that NSF probably will only fund GPU machines in the future. Care to comment?

A. I think that's too limiting. I think GPUs indeed are going to have a tremendous capability in the future but there's lots of other things coming up as well. For example, there's a lot of press about clouds, and clouds will indeed probably play a significant role, but we need to understand what can they do better than what exists today.

What we're trying to do with ACI is also look at the distributed nature of the compute and data side. For example, the increase in distributed capabilities coming from sensors, smart phones, tablets and laptops are already dramatically changing the computational world and will continue to do so in the future. With ACI, we need to look at the whole range of capabilities emerging, not just the very high end, nor just with GPUs. What do we need to do to best address the science, and most effectively increase productivity and national competitiveness?

Q. So how are we going to fund all of this? Will it still be on the five-year funding cycles?

A. Well some hardware is going to still be funded on a five-year cycle, but what's going to happen—and you're already seeing it on the data and software side—is more focus on multiyear programs and on sustainability. It is clear that while hardware may be obsolete in three or four years, software stays around for 10-20 years and data is forever. So we are looking at programs to address the much longer timeframes of these activities, what it means, and how to get there.

It is also important to note that ACI is not just an OCI activity; it involves all of NSF. For example, CISE (Directorate for Computer & Information Science & Engineering) will be announcing significant foundational research efforts in strategy number one; MPS (Directorate for Mathematics, Physical Sciences) will work to develop new mathematical algorithms. These will be multiyear programs.

We need to start looking at longer-term programs to be able to start focusing on those sorts of activities. A strict model of fixed three-to five-year programs is not adequate; the issues are much more nuanced and complex.

Q. NSF starting down a new path?

A. Exactly. We put down [in ACI] what we think are the issues. This includes a focus on enabling and facilitating foundational research in computational and data-enabled science and engineering, and broadening the use and applications to all disciplines.

Q. Anything else you want people to know?

A. If it's not clear, we're trying to talk about ways to significantly broaden and extend computational capabilities not only to all the disciplines, but also make it available to everyone. How do we provide computational access and expertise to the entire science community? What should NSF be doing for those communities that need enabling software, algorithms, and data rather than just large computational resources? In essence, how do we democratize computational and data-enabled science so that it permeates science and education?

Q. And doesn't that benefit all of us?

A. Yes. And that is what it is about. Plus, I'd point out, it's also a national competitiveness issue. Being able to effectively use computational tools, methods, and resources across the entire scientific spectrum is one of the most important things we should be doing to move forward.

You can download the "Advanced Computing Infrastructure: Vision and Strategic Plan" and the "Cyberinfrastructure Framework for 21st Century Science, Engineering and Education" at the NSF Office of Cyberinfrastructure website: www.nsf.gov/dir/index.jsp?org=OCI

A MOVING QUESTION

LOOKING FOR PLANETS SIMILAR TO EARTH IS "THE Holy Grail of planetary and extrasolar planetary studies," says Indiana University astronomer Scott Michael. That's why his group's discoveries concerning the movement of gas giant planets in protoplanetary disks are so significant. Improved understanding of the dynamics of the planets he studies may help others predict stable orbits of terrestrial-type planets.

Extrasolar planets, also called exoplanets, are planets outside of our solar system. Up until about 1995 there had been no exoplanet discoveries. The recent renewed interest in the formation of gas giant planets is due to the fact that over the past 15 years, says Michael, astronomers have discovered more than 500 stellar systems with planets. Some of those systems have multiple planets, raising the number of discovered exoplanets to over 700. Most of those exoplanets are gas giant planets.

Michael studies the migration of massive planets in protoplanetary disks. He first used NCSA's Cobalt and Ember and now uses the Pittsburgh Supercomputing Center's Blacklight. A protoplanetary disk is a gaseous nebula around a newly forming star called a protostar, also known as young stellar objects. Protoplanetary disks are flattened nebulae and have a flared disk shape.

"These are observationally confirmed objects; they have been seen. We're trying to study how a gas giant planet forms in such a nebula," Michael says. "By understanding whether they form at the location that we observe them or whether they move, we can then start looking into the planetary dynamics of an entire system."

Which is where the planetary Holy Grail comes in. Planetary astronomers believe that if they can find a planet similar to Earth's by Barbara Jewett

NCSA and XSEDE aid an Indiana University researcher studying the migration of massive planets in protoplanetary disks. Understanding how planets shift is one of the first steps in the quest to discover other life in the universe.

size and composition in orbit at an equivalent distance from a system's central star as the sun is from the earth, that it might be a habitable planet. Better understanding of planetary dynamics may yield the information they seek.

Planet formation

It's fairly well understood how terrestrial, rocky-type, planets form. Dust, rocks and boulders are in the nebula, and after the gas has partially dissipated the leftover bits collide, sticking together to build up a rocky planet. Planets like Earth, Mars, Mercury, and Venus. But the gas giant planets—those like Jupiter and Saturn—have a large fraction of their composition made up of hydrogen and helium gas, which astronomers believe must have come from a gaseous nebula.

"The rub in the whole thing is that those planets need to form in the time that the gas is around, before the nebula dissipates. Which is a fairly short timescale when it comes to astronomical time," says Michael.

Currently there are two theories as to how that can happen, he notes. One is called core accretion, also known as nucleated instability. In this theory a rocky core builds up through particles colliding and sticking together until it reaches a critical mass and starts to accrete the surrounding gas. Eventually it reaches the runaway growth phase where it very rapidly accretes all of the gas surrounding it within its gravitational influence. The result is a gas giant planet.

"There's a lot of issues with that theory, but one of the bigger ones is it takes quite a long time for that process to occur. And because the nebula doesn't tend to stay around for a very long time in astronomical timescales, that presents an issue with the timing," says Michael.

The other theory is called gravitational instability. That's the theory Michael's group studies.

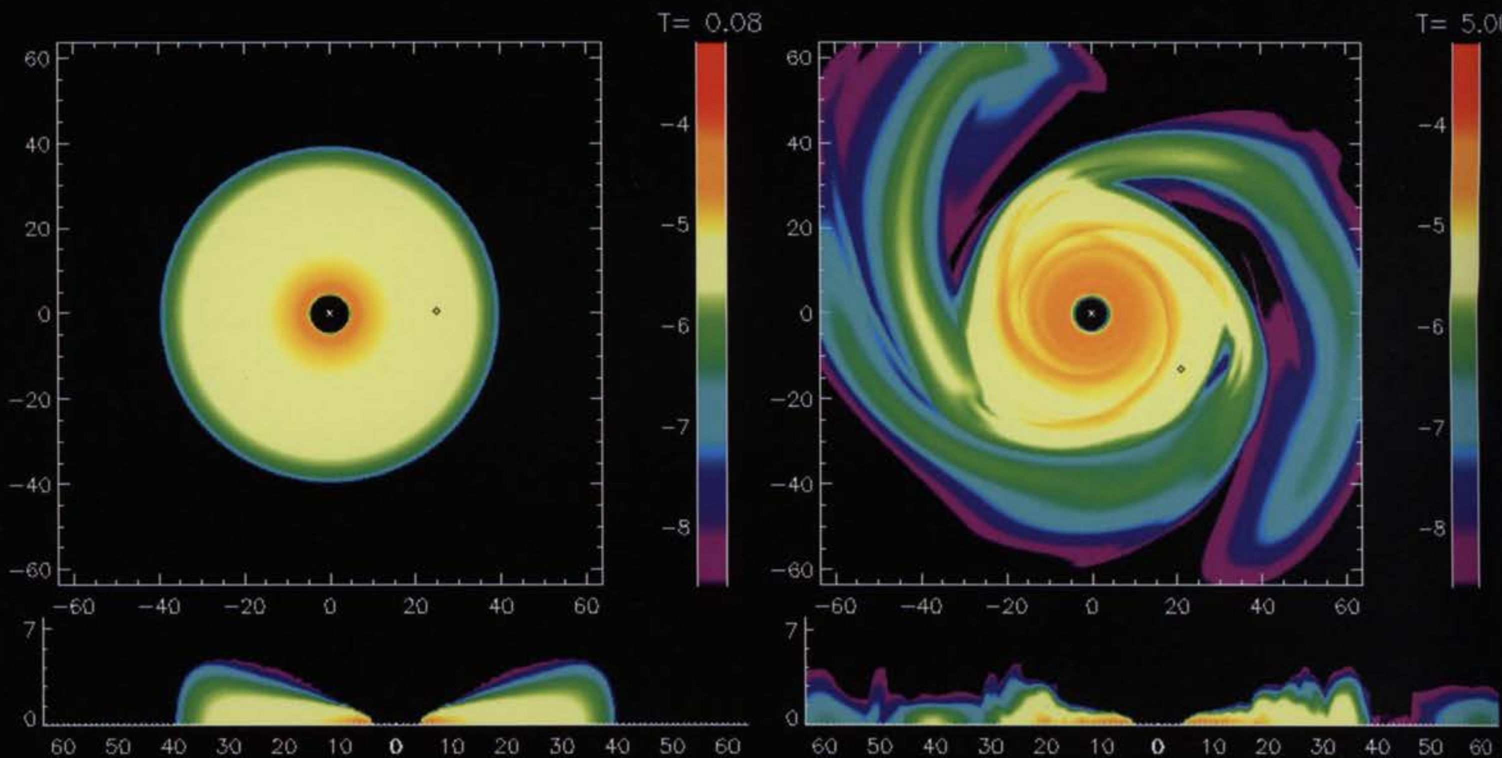
If the conditions are right in the protoplanetary disk—a high enough gas density and a low enough temperature—the disk will undergo a condition called gravitational instability. The force of gravity collapses parts of the disk to form very dense structures, perhaps collapsing to form a spherical object that would shrink down until it became a gas giant planet. Michael's team conducts simulations to look at the physical mechanisms that can affect when the gravitational instability might be active in a disk, how efficient those instabilities might be at actually forming gas giant planets, and under what conditions instability causes gas giant planet formation.

One of the benefits of the gravitational instability theory, says Michael, is that when it forms planets it forms them relatively quickly—within a few thousand to 10,000 years, well within the lifetime of the nebula. To form the gas giant planets in the core accretion model takes millions of years.

Planet migration

Scientists believe that, once formed, a planet in the nebula's protoplanetary disk will very rapidly move toward the central star, and may even be accreted on to the central star.

"Part of that theory fits nicely, as many of the extrasolar planets first discovered were found very close to their central star," explains Michael. "So the fact that planets rapidly migrate inward creates the question: What stops the planets from continuing to migrate inward and being accreted on to the central star? What mechanism could be responsible for these 'hot Jupiters'?"



ABOVE: Midplane and meridional densities in logarithmic scale for several different times in a protoplanetary disk simulation. This simulation follows the migration of 0.3 Jupiter mass planet inserted in an equilibrium disk at 25 AU. The axes have units of AU and the time is given in outer rotation periods or ORPs (one ORP = 180 years) in

The team looked at gravitational instabilities as a possible mechanism of altering that pattern of migration and causing planets to stop. In a variety of different simulations some of the planets they put in the protoplanetary disk began to migrate inward very rapidly but then stopped at a certain radius. They were able to correlate the place that they stopped with the gravitational instabilities tapering off as well.

"This would explain why planets halt their inward migration and could explain those gas giant planets found close to the central star," Michael says.

But the team had another interesting discovery as well. When they put planets into a gravitationally unstable disk, in some cases the planets actually migrated outward.

That fits with exoplanet discoveries. As the observational community matured and looked for different techniques to discover exoplanets, says Michael, they began to find gas giant planets at a variety of radii. In fact, he notes, there are planets discovered at very large distances from their central star.

In a typical protplanetary disk, the surface density of the nebula decreases as you go farther and farther away from the star. At very large radii the gas probably isn't dense enough to form gas giant planets via core accretion, he says. So, his team's finding that gravitational instability can in fact cause planets to migrate either inward or out away from their central star is significant, as it might help explain the vast range of separations from the central star at which gas giant planets are found. Initial results of Michael's studies have been published in the *Astrophysical Journal Letters.*.

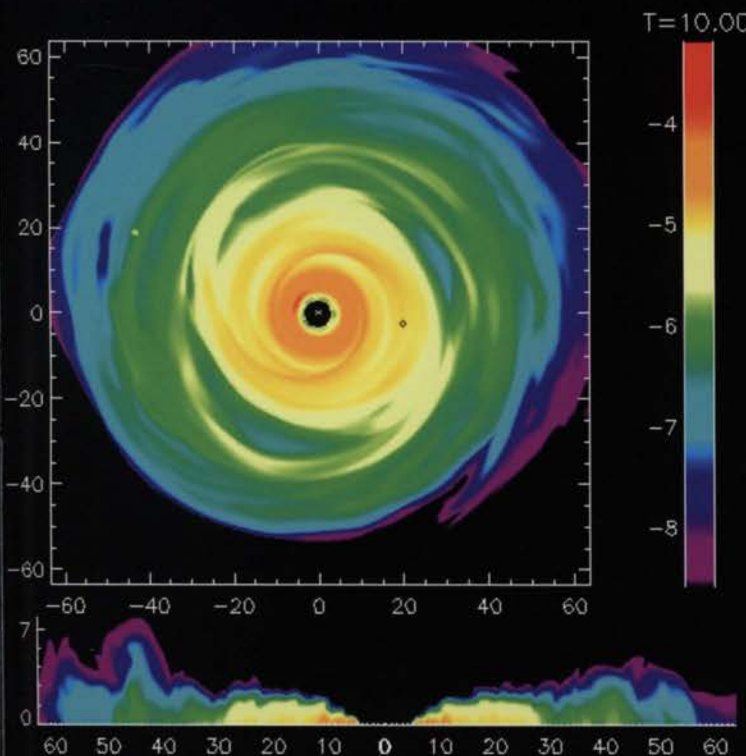
Simulating the systems

The team does simulations using a code with a fixed cylindrical grid, meaning there are a certain number of grid cells in each of the directions r, Ø and z. The amount of physical time simulated for each time step of the simulation is limited by the cell that has the most rapidly moving gases. When they get to a part of the nebula in the simulation that becomes very dense or very hot, it will have very rapidly moving gas and that causes the time steps to become very short in physical time.

Michael says his team developed a very nice workflow using first NCSA's now-retired Cobalt, then Ember. Since their code can only run on machines that have shared memory architecture, they moved to PSC's Blacklight when Ember did not transfer into the XSEDE resource system last year.

"One of the big issues when we first started using the supercomputer was transferring the data from the machine to our resources at Indiana in order to do the analysis. The solution we were able to employ was a piece of technology we came up with at IU called the Data Capacitor on the wide-area network, or DC-WAN," says Michael.

With the DC-WAN, the team can run the simulation and have the output data directly written at their university. They can even



the upper right of each panel. The black diamond in each panel indicates the location of the planet, and the white X in the middle is the location of the central star.

PROJECT AT A GLANCE

TEAM MEMBERS

Scott Michael (Indiana University)
Richard H. Durisen (Indiana University)
Thomas Y. Steiman-Cameron (Indiana University)
Aaron C. Boley (University of Florida)
Megan K. Pickett (Lawrence University)

FUNDING

NASA

FOR MORE INFORMATION:

http://hydro.astro.indiana.edu

ACCESS ONLINE:

www.ncsa.illinois.edu/News/Stories/disks

do analysis in real time to see what is going on as the simulation progressed. That workflow works really well, says Michael.

"We ran simulations on Cobalt and Ember using between 64-256 cores, data written directly to DC-WAN here at IU and we did our analysis as soon as it arrived," says Michael. "Working with the folks at NCSA was wonderful. They were so helpful and willing to work with us on mounting the DC-WAN file system on the machine. It was really a sad day the day they took Ember off of XSEDE as it denied us a really valuable and useful resource."

But he was able to move the hundreds of thousands of core-hours allocated to his project on Ember to the Blacklight machine at Pittsburgh, and his team established a similar workflow there.

"We were lucky that we were able to get almost all of our allocation transferred over to Pittsburgh. But this was a fairly major disruption for us, as it took almost three months to get everything set up on Blacklight so we were back in full production mode," he notes.

Typically the team will perform between 1 to 4 million timesteps over a grid of 16 million computational cells. That will take approximately a month or so of continuous wall time, he says.

"In practice, it takes about two months because you have to checkpoint and restart systems on these big HPC machines. The longest stretch you can usually run is four to five days. So we'll run for a few days, we'll checkpoint, then we'll resubmit the job. We'll take a look at the results from our analysis on DC-WAN, resubmit the job, wait a few days in the queue due to Blacklight's demand, then it will run another four to five days. And we just keep on doing that. What we've started doing when we want to explore parameter space, say we want to compare the migration rates of planets of different mass, is to run several simulations at the same time. That way we can get three or four simulations done in a two-month time period," says Michael.

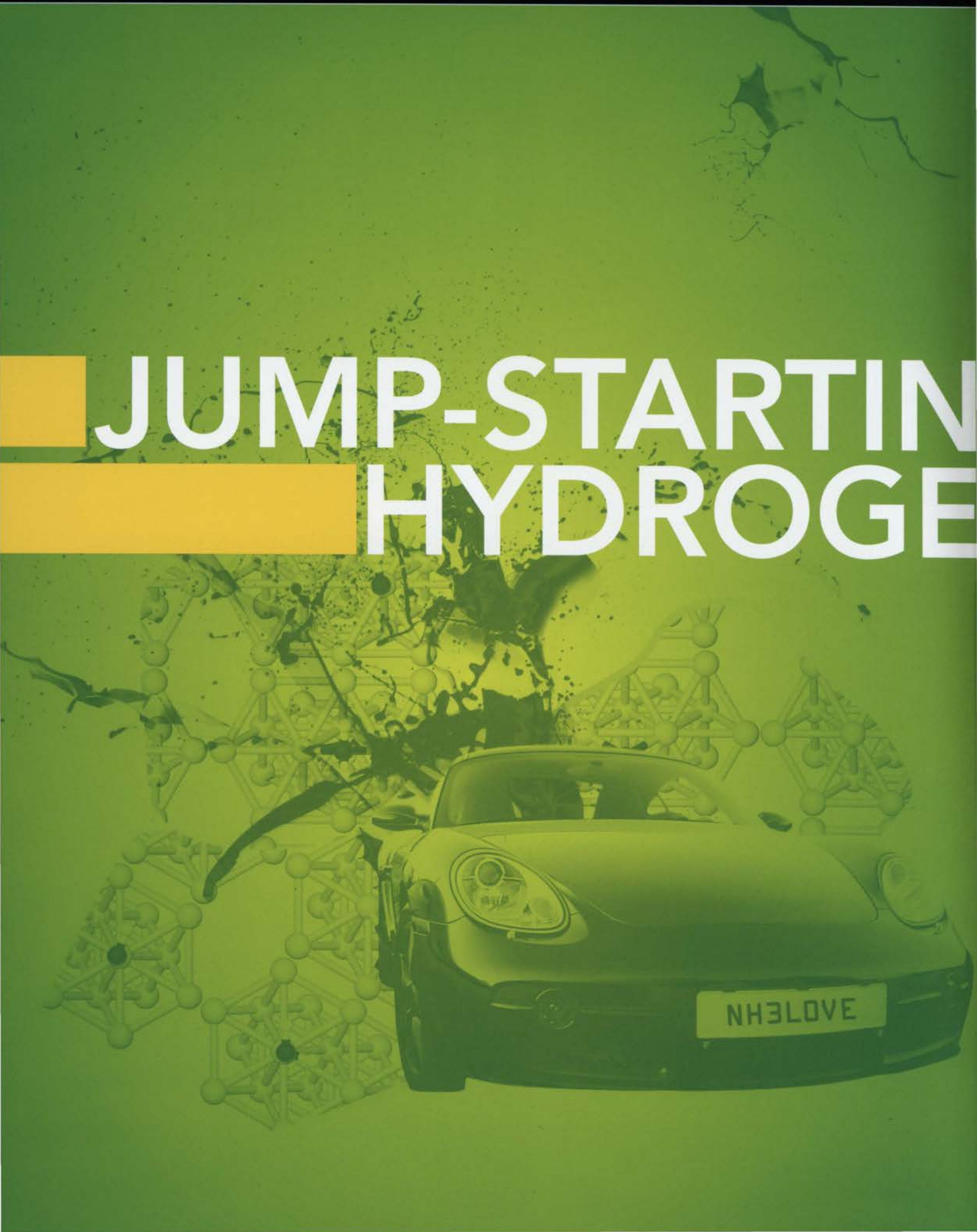
Next phase

Almost all of the team's studies have been focused on nebulae about the size of our solar system, about 40 AU in radius. The unit AU stands for astronomical unit, defined as the distance between the Earth and sun; the distance from the sun to Pluto is about 40 AU.

There have been observed protoplanetary disks that are 100, 200, even up to 400 AU in extent, says Michael, and members of the team are now beginning to simulate those.

"It turns out that even though there a lower surface density at those larger radii, because it is so much colder being so distant from the central star they actually have a much higher likelihood of becoming gravitationally unstable," he explains. "So we're trying to understand if a gas giant planet could form at very large radii given gravitational instability. Then this ties into the whole question of migration. If gas giant planets were to form at 100 or 200 AU, how could they move much closer to the central star and not get accreted on to the central star? It will be very interesting to see what the results show." □

JUMP-STARTIN HYDROGE



by J. William Bell

With computing resources from XSEDE, engineers at Ohio University explore ammonia as a source of hydrogen for tomorrow's fuel cells.

G THE
N ECONOMY

F

OR THE FUEL CELL INDUSTRY, AMMONIA HAS all kinds of things going for it. It's abundant—the U.S. Geological Survey reports that more than 130 million metric tons were produced worldwide in 2009. The pipelines and trucks to ship ammonia around the country are already in place. It releases nothing but nitrogen and water vapor into the atmosphere when used for fuel cells. And, according to the Iowa Energy Center, it's at least as safe as gasoline when used as a transportation fuel.

But many have considered ammonia a dead end. It requires a lot of heat to break the hydrogen, which is what most fuel cells generate their energy with, out of the ammonia. By-products of this process can also foul the fuel cells and reduce their efficiency.

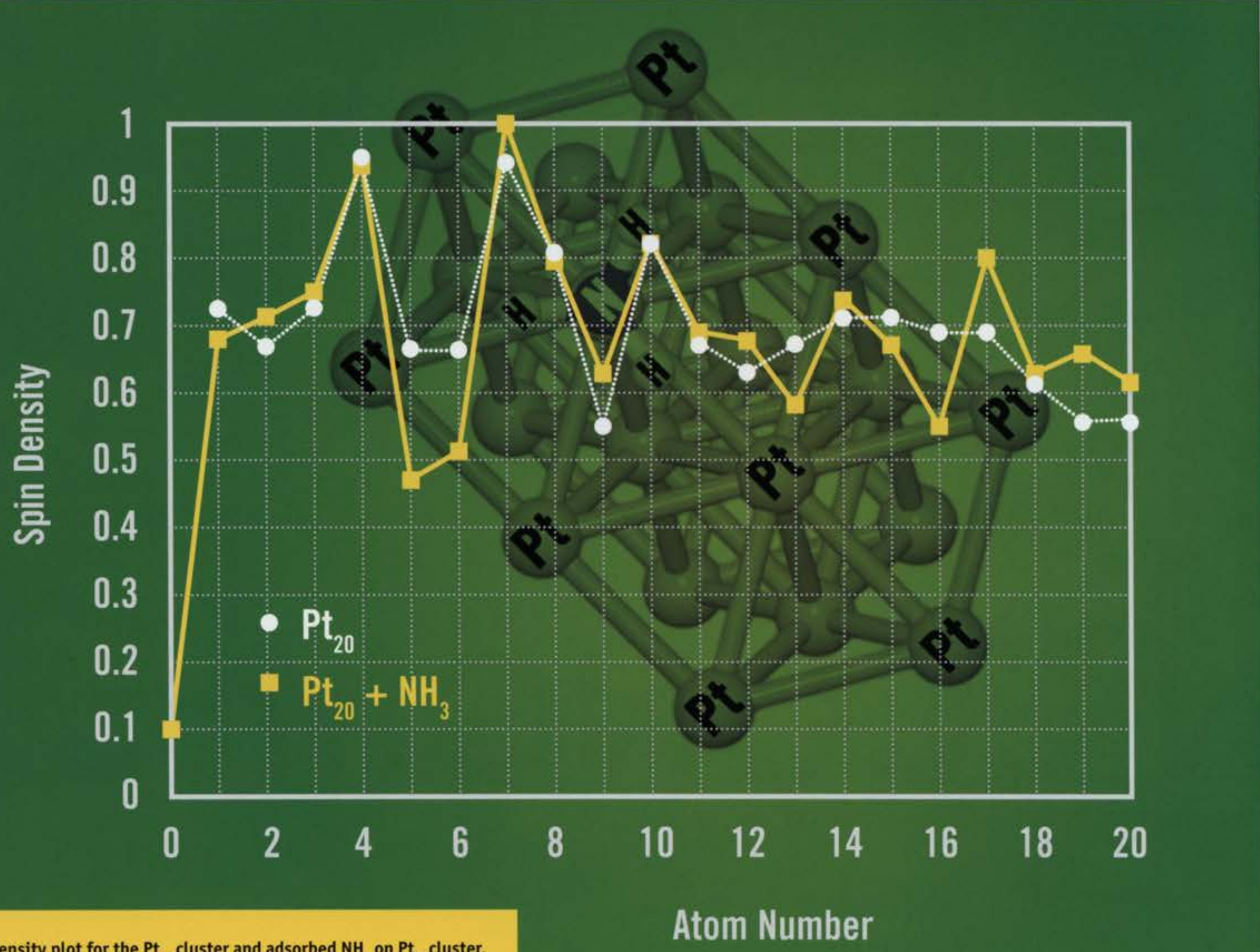
With XSEDE resources at NCSA and the Pittsburgh Supercomputing Center, a team of engineers at Ohio University's Fritz J. and Dolores H. Russ College of Engineering and Technology is jump-starting the prospects of ammonia-based fuel cells for cars and other applications. XSEDE is a collection of advanced digital resources around the country, a single virtual system that scientists can use to interactively share computing resources, data, and expertise. XSEDE is led by NCSA.

The team sees ammonia-based fuel cells as a way of shifting the United States, and the world, away from petroleum and other fossil fuels and toward a hydrogen economy. That economy would require hydrogen that is "cheap and produced at the point of use. Ammonia can provide that," according to Damilola Daramola, a post-doctoral research associate working with Russ professor Botte.

Matt Simmons, who before his death was a member of the Council on Foreign Relations, agreed. Simmons was a long-time advocate for solutions to "peak oil," the point at which the world begins consuming more oil than it can extract and refine.

In a 2010 report, he said "NH_3 [or ammonia] is the only realistic energy solution that makes sense."

The Ohio team's results have been published in *Computational and Theoretical Chemistry*. Their approach to breaking the ammonia down for use in fuel cells has also been licensed by the company E3 Clean Technologies for possible use in cleaning up wastewater. (Ammonia, by way of urine, is a major component of wastewater that has to be removed.)



Spin density plot for the Pt_{20} cluster and adsorbed NH_3 on Pt_{20} cluster, showing minimal change in spin upon NH_3 adsorption.

1.5 gets you 33

An ammonia-based fuel cell relies, like the vast majority of fuel cell designs, on hydrogen. Using a small electric current and a catalyst to drive the chemical reaction, the hydrogen atoms are split into protons and electrons. The protons pass through a selective membrane, while the electrons are forced through a separate external circuit to generate electricity. On the far side of the membrane, the electrons and protons combine with oxygen to form water. In the case of ammonia, a harmless gas like nitrogen is a by-product.

In the case of an ammonia-based fuel cell, however, the hydrogen is created by "cracking" the ammonia into its component parts. Unlike hydrogen, ammonia does not have to be transported under pressure, reducing the risk of explosions and other mishaps when carrying the fuel on board a car.

Older designs for cracking ammonia were inefficient.

However, Madhivanan Muthuvel, a research assistant professor in Botte's lab at Ohio University's Center for Electrochemical Engineering Research, told The Columbus Dispatch that the team's approach addresses that issue. 1.55 watt-hours of electric energy, using Botte's patented method, yields a gram of hydrogen from ammonia. That hydrogen can then produce 33 watt-hours of electricity from a fuel cell.

One kilogram of hydrogen produced this way using ammonia costs about 90 cents. To produce that kilogram of hydrogen using water runs about $7. And that 90-cent kilogram of hydrogen can deliver roughly the same amount of energy as a gallon of gas.

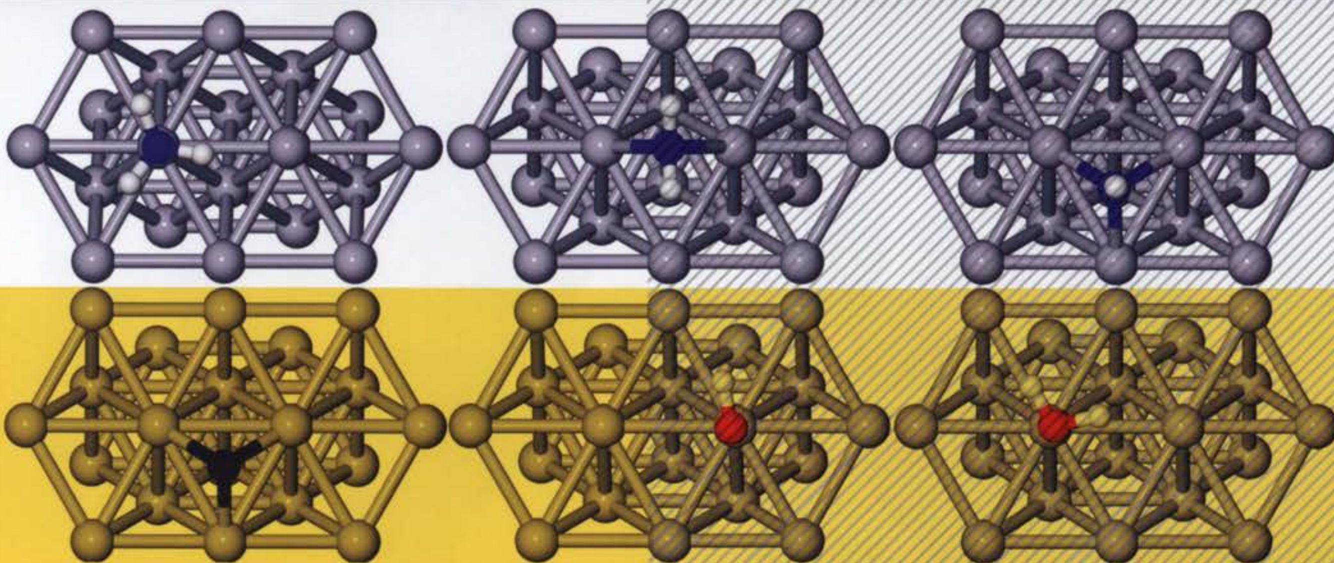
A Goldilocks reaction

The Botte team's work has an experimental element. They actually build "electrolyzers" that zap ammonia into its component parts. Those are used to test the overall efficiency of the system and how changes to the system—changing the elements used to make the catalyst that drives the reaction—impact that efficiency.

The computational element on XSEDE supercomputers is crucial as well.

"The computation strengthens our arguments by telling us why the reactions behave as they do," explains Daramola. Not to mention the fact that their computational experiments are much less expensive to run than the equivalent lab-based experiments would be.

The team models the molecules that form and are then further broken down in the process of sheering hydrogen from ammonia molecules. Specifically they consider the strength of the bond between these intermediates and the platinum catalyst that drives the



The species adhering to the surface of platinum during ammonia conversion occupy the special positions of high symmetry shown. These species are (clockwise from top left) NH_3, NH_2, NH, H_2O, OH and N.

interaction. They also look at the orientation of those molecules as they interact with the catalyst, as well as how much and in what ways individual atoms within the intermediates are moving.

They then rank the intermediates in terms of which are most likely to adhere to the platinum, which in turn tells them which are most likely to pollute the catalyst over time and impinge on how well it works.

Understanding these features allows the team to look for a Goldilocks reaction that produces the most hydrogen while degrading the catalyst as little as possible. It also means they can begin to explore platinum alloys—platinum mixed with elements like iridium or rhodium—that might be more efficient or cost less money.

"You need the intermediates to stick to the surface of the catalyst and become converted to the desired product. But you don't want to make it too sticky such that it adheres too strongly to the surface or not sticky enough such that the conversion will not occur," Daramola says.

The Ohio University team runs their simulations on the Blacklight system at the Pittsburgh Supercomputing Center and the Ember system at NCSA. □

PROJECT AT A GLANCE

TEAM MEMBERS

Gerardine Botte
Damilola Daramola
M. Muthuvel
Alex Miller
John Goettge
R. Palaniappan
Luis Diaz-Aldana
Dan Wang

FUNDING

U.S. Army Construction Engineering Research Laboratory

FOR MORE INFORMATION:

http://www.ohio.edu/ceer

ACCESS ONLINE

www.ncsa.illinois.edu/News/Stories/ammonia

IMAGE

MINERS

BEHIND THE SCENES OF NCSA

THE INFORMATION AGE WE LIVE IN EXISTS ON TOP OF a large amount of digital data. Thanks to the web we have access to this data. Thanks to search engines that index this data we can efficiently search it. What many may not realize is that this is only a fraction of the data that could be at our disposal. For millennia humans have recorded information on media such as paper. This information exists today in libraries and archives throughout the world with much of it not truly being a part of the information age.

Most text we find on the web was entered digitally by a human who typed it on a keyboard. To go back and do this for all the archived records created before the invention of the computer is an enormous and costly task. A much more realistic and practical means of digitizing this information is to scan it as a digital image. Digitizing old data in this manner, however, only solves one of our two problems, that of providing access to the data, not searchable access. This problem is exemplified by work being undertaken for the 1940s Census data release which will be in the form of 3.8 million JPEG images (see pg 20). Anyone can download this data and look at it. But then what? Can one find anything by looking through these millions of images containing billions of individual entries?

Though still a relatively young field, it is problems like these where the field of computer vision can offer some solutions. The burgeoning amounts of raw image data many organizations, agencies, and companies must deal with today have made even imperfect solutions highly desirable if they allow one to make some sense of large image collections. The Image, Spatial, and Data Analysis Group (ISDA), formed nearly a decade ago by Peter Bajcsy and currently led by Kenton McHenry, conducts research and development involving image and video data. Sifting through the state of the art in computer vision the ISDA group applies research to real world problems while striving to build robust multipurpose software to serve a variety of community needs. The group works on providing a framework for searchable access to large collections of handwritten forms, creating suites of semantic keyword extractors from image content, developing tools to aid in the study of group interactions in collections of videos, developing tools to aid in the identification of authorship in old manuscripts, and stitching together large macro images from collections of overlapping photographs. With decades of paper archives in the United States alone coupled with low-cost ubiquitous digital cameras that we all have today, efforts to unlock large collections of image data is becoming a necessity while also being very much a computationally intensive supercomputing problem. □

ISDA team members (clockwise, from top left): Kenton McHenry, Devin Bonnie, Luigi Marini, Michal Ondrejcek, Rob Kooper, Liana Diesendruck

TIMING IS EVERYTHING

By Barbara Jewett

NCSA's GPU machines allow researchers to improve the fluid dynamics codes used for rotorcraft design.

A visualization of the wakes coming off a helicopter rotor. The simulation was conducted with a new coupled Eulerian and vortex particle method code. The visualization was then created using FieldView. Image courtesy of Intelligent Light. ©2012, Intelligent Light.

THERE'S AN OLD JOKE THAT HAS A FLIGHT INSTRUCTOR telling his student takeoff is easy, it's landing that's hard. Helicopter and tiltrotor pilots know that landing is even more difficult if you're battling the wakes generated by the rotors.

Unfortunately, rotor wake is a major problem in rotorcraft. That's why Earl P.N. Duque, manager of the Applied Research Group (ARG) at Intelligent Light, a New Jersey-based software company specializing in computational fluid dynamics (CFD) and makers of the CFD post-processing software FieldView, together with Christopher P. Stone, founder of Computational Science & Engineering LLC, based in Chicago, used NCSA's Forge and now-retired Lincoln to develop and test a simulation application that more accurately predicts rotor wake interactions. ARG was a participant in a five-year project working on CFD research for rotorcraft as part of the Army's Vertical Lift Research Center of Excellence based at Penn State.

Rotorcraft refers to anything that has vertical lift from a rotor—unmanned air vehicles (UAV), helicopters, or the newer tiltrotors like the V-22 Osprey. The Osprey has a three-bladed proprotor, a turboprop engine, and a transmission nacelle mounted on each wingtip. For takeoff and landing, it typically operates as a helicopter with the nacelles vertical and rotors in a horizontal position. Once airborne, the nacelles rotate forward 90 degrees in as little as 12 seconds, converting the V-22 to a more fuel-efficient, higher-speed turboprop airplane. The U.S. Marine Corps began using Ospreys in 2007 and the Air Force in 2009. They have been deployed in Iraq, Afghanistan, and Libya in both combat and rescue operations.

The challenge

"The big problem with helicopters is the interaction between the vortex wake coming off the rotor blades and how those vortex structures affect the lift behavior and drag behavior of the rotor blades themselves," explains Duque.

He notes that the disturbed air may further change as it passes over the fuselage of the vehicle, creating new wake structures. This can impact the vehicle's tail, causing control problems in the tail area.

The traditional way of developing new helicopters is to design one, build it, and then fly it to test it. It's not uncommon during this demonstration and validation phase to discover significant wake problems.

"One of the big problems whenever there is a new design for a helicopter," Duque says, "is that inevitably a tail rotor shake happens. That means the wakes come off the rotor hub or the rotor blades themselves, hit the tail, and cause problems with control of the vehicle. Problems like the pilot can't land."

To alleviate tail shake requires extensive wind tunnel testing and flight tests, which are costly. Thus the Army is sponsoring research to develop the software capability to predict tail shake problems in helicopters and UAV during the design stage. Being able to accurately model the wakes and predict tail shake will dramatically reduce the number of prototypes built and the number of wind tunnel or flight tests required.

A solution

For their project, says Stone, they took standard grid-based finite-difference (Eulerian) CFD software and fused that with a particle-based method that tracks vorticity, the vortex particle method. Vortex particles are chunks of fluid that are free to move around to represent the motion of the fluid, as opposed to the grid-based system where one tries to solve the equation at each static grid point.

"We came up with the idea to combine these two disparate simulation methods that we theorized would enable us to better predict the wake interactions. 'Better' meaning more cost effectively and also more accurately," Stone says. "The grid-based methods have been in use for a very long time because the software was already there, and the numerical methods are pretty good for modeling the flow close to the fuselage body or the rotor wing. But farther away from the body the grid-based methods are not very accurate."

If the CFD used is not accurate enough, Stone explains, the vortex that's striking the tail rotor won't have the right physics thus causing the tail rotor shake. With the particle method, he and Duque were hoping to more accurately predict how a vortex would flow back and strike the tail rotor, providing improved predictions of the rotorcraft's handling.

One of the biggest challenges with developing the new code became apparent early in the process: the particle-based method turned out to be extremely expensive. It was not how the program was written, says Stone, "it was just mathematically a very expensive way of doing the calculations." But the GPU computing revolution was beginning, and the technology helped him quickly overcome the cost issues.

"The year we started this project, GPUs came into use. So we were able to make use of GPU computing to really get this project going. We started using Lincoln right when it was deployed. Now we use Forge," says Stone. "We were able to redesign our particle codes to use GPUs and honestly, without those GPU capabilities, we would not be able to do the simulations that we're doing."

Stone notes performance was not just a couple factors faster, there was an order of magnitude difference. But more importantly, he says, is that with access to Lincoln and Forge he and Duque were able to attempt simulations that were not previously possible.

Once the codes were modified for GPUs, they focused their attention on fundamental research, looking at the algorithms and how to couple the grid-based code with the particle code. They looked at fluid flow against a static wing, trying to predict some simple flows in order to validate the research. They then progressed to looking at loads on the rotor and fuselage in real-world flight conditions and then used Intelligent Light's FieldView software to visualize the predicted rotor wakes interacting with a helicopter fuselage as shown in the images accompanying this article.

The past nine months they've been working on this with the ROtor-Body INteraction (ROBIN) fuselage shape, which is used by Army and NASA rotorcraft groups for code and measurement validation. Their original theory that the particle method is superior to predict the rotor wake far downstream was upheld with this project, notes Stone. But they are still working on the validation.

Another important issue when it comes to interaction of the vortices and the rotors, says Duque, is the audible noise generated from the helicopters themselves. A blade hitting a vortex, as well as the vortex generated by the blade ahead of itself, can cause an undesirable noise character the Army would like to

eliminate. Currently there is no software capable of modeling those interactions.

"It's research," says Duque. "We've shown the methodology works, but there is still work to be done."

Code potential

Although the team's work was focused on rotor-wake interaction on the same vehicle, Duque says they briefly looked at predicting wakes multiple rotor lengths downstream. They also looked at other potential applications for the code, such as wind energy projects where the wake interaction between turbines is important.

One of the interesting tangential outcomes from this research, notes Duque, is that while there are other people doing similar types of CFD work, in the rotorcraft world Stone was the first to utilize GPUs in the coupled Eulerian and vortex particle method codes.

So in addition to being cutting-edge on the method, they are also leading the way with the technology employed. Duque says he thinks that the knowledge they gained of how to use GPUs was one of the most beneficial aspects of the project. In fact, he says, he has made use of GPUs on other projects.

"We turned out a pretty open-ended research project," Stone says. "We came up with some good methods but it is not yet 100 percent. The main problem is that the grid-based codes tend to use a compressible flow, and the particle methods use an incompressible. So how to join those two methods turned out be extremely complicated to get the physics right. We've come up with a good way of doing that, which gives decent results. But it's still an open question of the exact mathematical way of doing it right. With access to GPU resources, though, we'll get there."

PROJECT AT A GLANCE

TEAM MEMBERS

Earl P.N. Duque
Christopher P. Stone

FUNDING

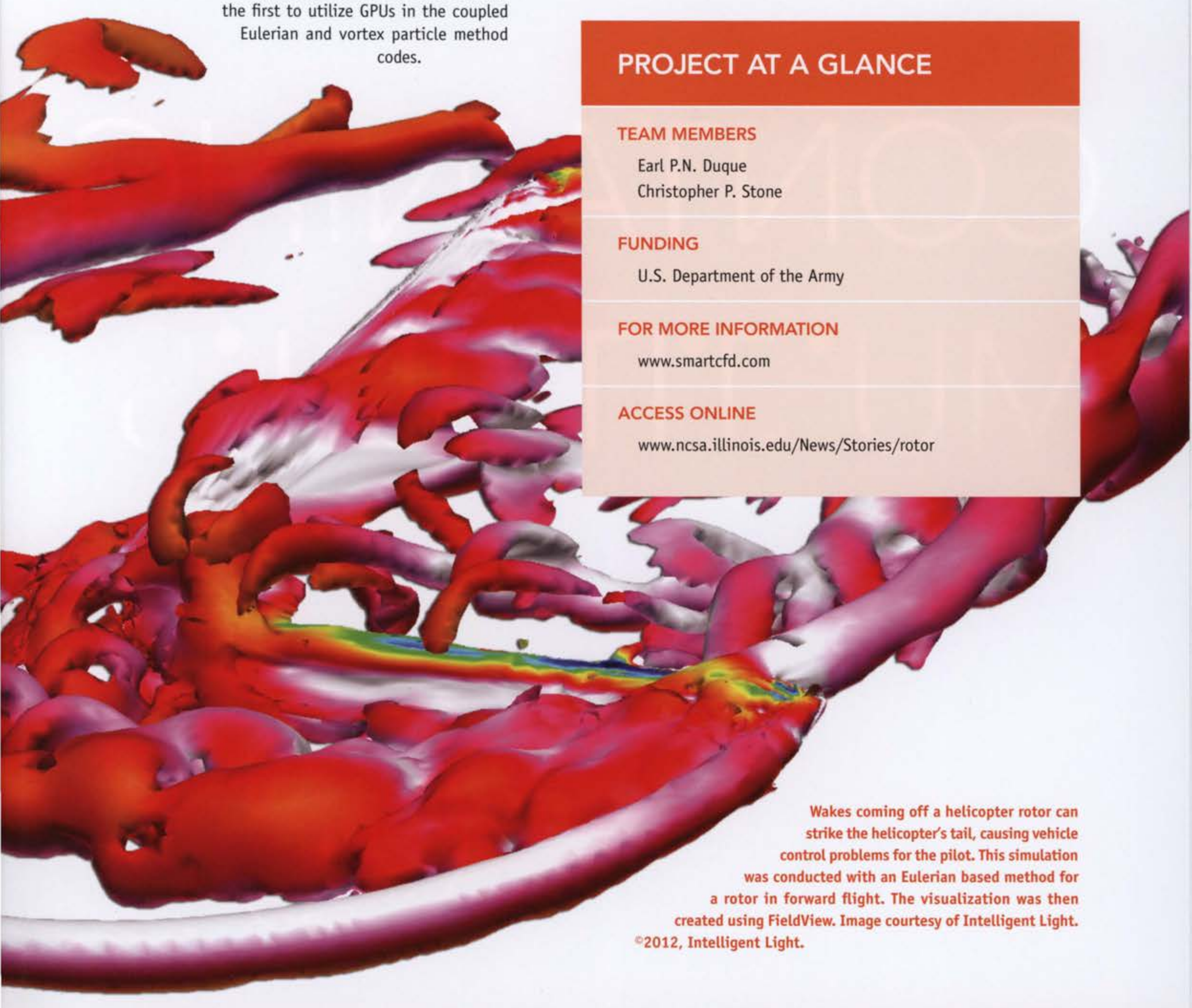
U.S. Department of the Army

FOR MORE INFORMATION

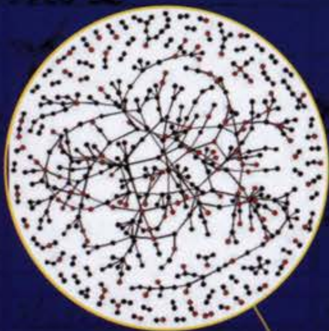
www.smartcfd.com

ACCESS ONLINE

www.ncsa.illinois.edu/News/Stories/rotor



Wakes coming off a helicopter rotor can strike the helicopter's tail, causing vehicle control problems for the pilot. This simulation was conducted with an Eulerian based method for a rotor in forward flight. The visualization was then created using FieldView. Image courtesy of Intelligent Light. ©2012, Intelligent Light.



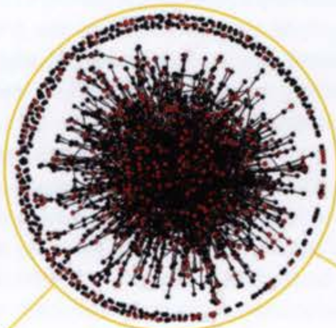
Players trade desirable in-game items with others or sell them in exchange for in-game currency. Face-to-face item exchange reveals individual connections between the players: the two parties need to know each other's names and meet physically in the game to finish the transaction. Based on the players' face-to-face trade, the VWE researchers constructed player trade relation as a binary and undirected tie if two players exchanged at least one item during the sample time period.

CONTAINING MULTITUDES

Each network graph illustrates the relationships among 3,140 unique players who played EverQuest II between Aug. 25 and Aug. 31, 2006.

by Trish Barker

NCSA is helping humanities and social science scholars analyze troves of data about worlds both real and virtual, shedding light on human behavior.



In the world of EverQuest II, players can form cooperative "groups" in order to complete tasks. If two players pair up and earn experience points in combat activities, such as fighting monsters, the VWE researchers construct a partner relations between them. In this instance, the researchers considered only pairs because this form of collaboration represents a stronger dyadic relation than the interactions in bigger groups. The extracted relations are used to construct a binary and undirected graph where nodes represent individual players and edges are their partnership.

This graph represents each time two players chat with one another in game.

ACROSS THE SPECTRUM, DATA HAS GOTTEN BIG. "If you look at the trend, databases are getting bigger and bigger," says NCSA database architect Dora Cai. While 50 gigabytes would have been considered a large database not that long ago, "now we're talking about terabytes and hundreds of terabytes and even petabytes."

The Virtual Worlds Exploratorium and an ongoing census analysis project are two examples of data-intensive research in the humanities that show how NCSA's infrastructure and staff can help researchers address the challenges of big data.

Interacting in Virtual Worlds

Millions of people around the world play massively multi-player online role-playing games. And as they play, their every action—each time they fight a dragon, buy or sell armor, talk to another player—is logged by the game, creating a wealth of information about how people interact in these "virtual worlds."

Several years ago, Sony approached researcher Dmitri Williams, then at the University of Illinois and now at the University of Southern California, to see if he could use data gathered from EverQuest II to determine which players were likely to leave the game (and therefore stop paying to play). Williams was also interested in questions about whether in-game behavior correlated with behavior in the real world. Will someone with a violent, aggressive game character be more violent or aggressive in the real world, for example?

Williams teamed with co-principal investigators Marshall Scott Poole (Illinois), Nosh Contractor (Northwestern), and computer scientist Jaideep Srivastava (Minnesota) to investigate a massive collection of game log data from EverQuest II and other games—Dragon's Nest and Chevalier's Romance, which are popular in China, and Denmark-based EVE Online. They call their collaboration the Virtual Worlds Exploratorium (VWE).

The researchers faced several challenges in working with these data:

- *Volume:* The logs added up to tens of terabytes.
- *Security:* The data needed to be kept secure and confidential.
- *Heterogeneity:* The data originated from different sources, was generated by different programs, and was in different formats and even different languages.

The data are housed at NCSA because "they have a lot of experience with large data and with making sure the data is securely handled," Contractor says. And the VWE team worked with NCSA database architect Dora Cai to create an organized database from the "messy" collection of log files.

If you aren't a data-focused researcher or computer scientist, you might miss the significance of that crucial step, but a collection of data isn't a useful database until it has been organized and structured and can be queried. On this project and many others, Cai was responsible for "architecting the solution so the database becomes a useful research tool," says Tim Cockerill, associate project director for XSEDE (the Extreme Science and Engineering Discovery Environment).

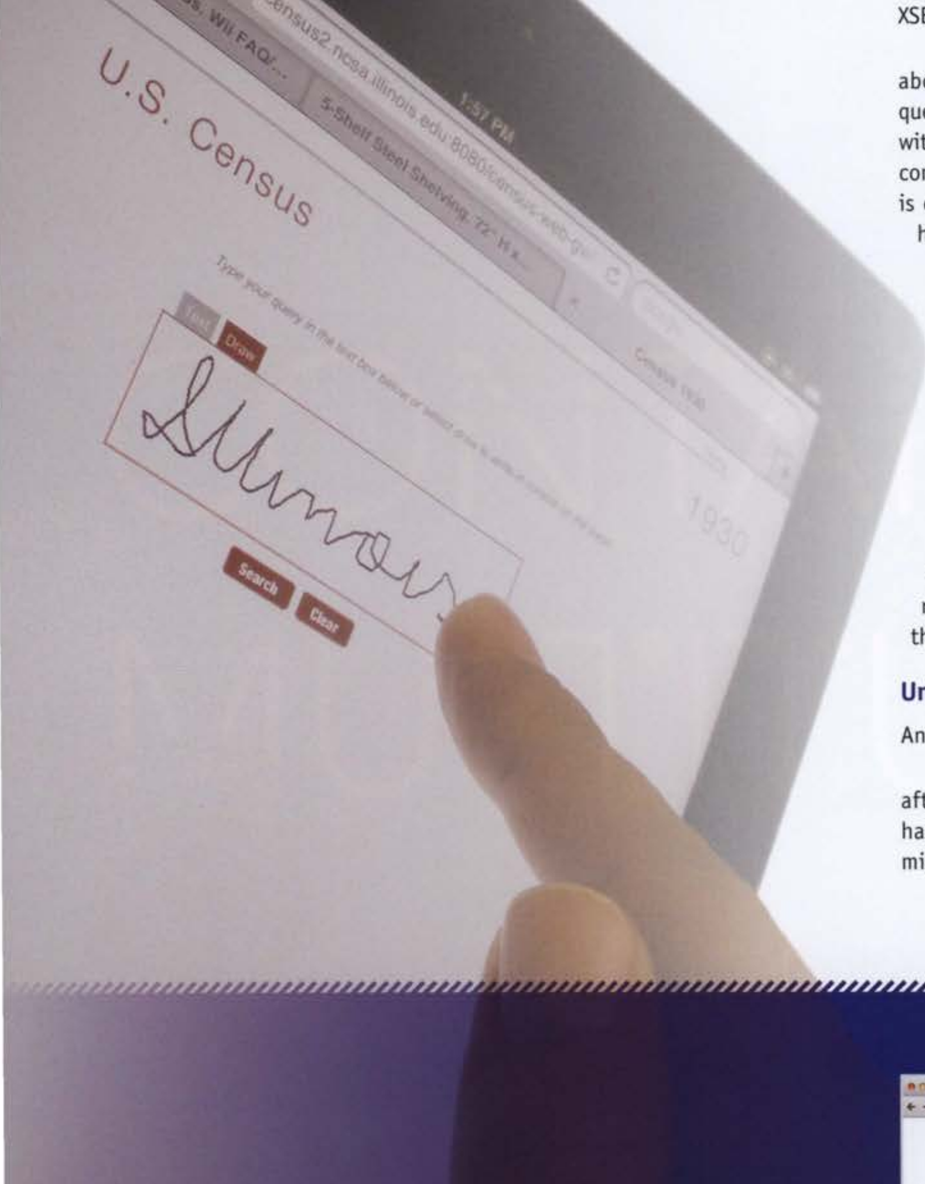
"We have all these great data, and we can ask loads of questions about interaction in the space," Contractor says. Some of those questions have addressed group formation (Why do people team up with one another in the game? Do groups form based on similarities, complimentary differences, proximity, etc.?) and leadership. The latter is one of the areas of interest to the Army and the Air Force, which have both provided funding for VWE projects. "This might be the best training ground for the kinds of leaders we will see tomorrow," Contractor says.

The researchers have also studied "illegal" transactions in which players sell currency, items, and even high-level characters to wealthier players who want the perks without putting in hours of game play to earn them. As games try to crack down on this behavior, the illicit sellers and buyers adopt new tricks to conceal their actions. One of Contractor's students, Brian Keegan, along with fellow student Muhammad Ahmad, found that the "illegal" networks in the game employ virtually identical strategies to those used by drug traffickers. The researchers also found that people who engage in illegal conduct in the game are more likely to have real-world criminal records.

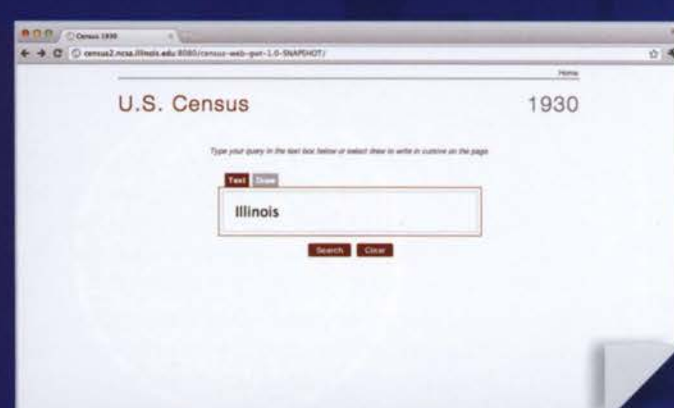
Unlocking secrets of the Census

Another NCSA big data project involves the real world.

A treasure trove of U.S. Census data is released to the public after remaining confidential for 70 years. The standard practice has been for the Census Bureau to create microfilm images of the millions of paper forms. Companies that cater to genealogy buffs,



FROM LEFT TO RIGHT A snapshot of the Census framework, showing how a query can be input as typed text (which will then be rendered to look like handwriting) or handwritten. The search for "Illinois" returns an accurate match.



like Ancestry.com, then hire thousands of people to spend months transcribing the microfilm so the data can be searched and sorted online.

But this April the detailed information on the more than 132 million people who lived in the United States in 1940 will be released in digital format. No more microfilm.

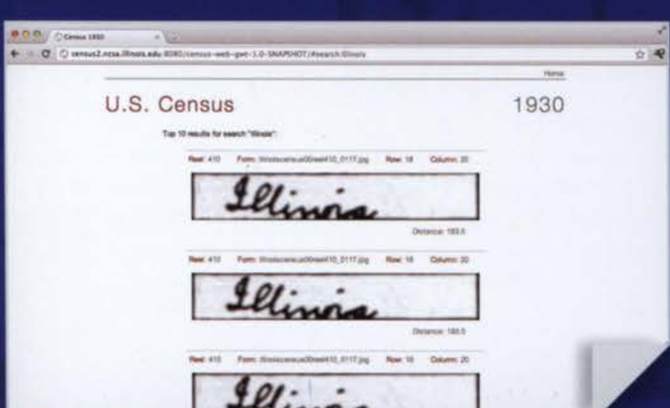
The Census Bureau would like to provide something more usable than 3.8 million JPEG images of census forms, but manual transcription is too expensive, and optical character recognition of the handwritten entries is not accurate enough. So NCSA's Image, Spatial, and Data Analysis group, led by Kenton McHenry, has been working for the past year on a prototype framework using content-based image retrieval to allow people to search the census form images directly. The project is supported by the National Archives and Records Administration.

The framework enables a user to input a handwritten query—either using a stylus or by typing a word that will be then rendered in a handwriting font—to search a database of images of handwritten text for potential matches. Using a computer vision technique known as word spotting, the top ranked results are returned.

While not all will be perfect matches, the system's users will help improve the results over time through a passive form of crowd sourcing. For instance, after searching for "Smith" a user isn't likely to click on results that are not "Smith." The query text entered by the user can be connected to the image results the user selected, allowing the image database to be slowly annotated. Over time, the validated matches can be returned to users rather than relying solely on the word spotting technique.

A significant amount of computation is required in order to pre-process the data to allow for the planned word spotting and passive crowd sourcing. The first step is to split the spreadsheet-like Census forms into individual data cells by finding the form lines and fitting a template over the images. Next, each extracted cell must be converted into a numerical feature vector that roughly represents the handwritten contents of that image. A word spotting technique compares the feature vector of the search query (such as a name, like Smith) to the feature vectors of the many, many cells, looking for similarities. To search all 70 billion cell images would be excessively time-consuming and computationally expensive, so a third step groups similar feature vectors and constructs a hierarchy on the data to narrow the search space and return results with reasonable speed.

The team is using an XSEDE start-up allocation to develop their system. An XSEDE Extended Collaborative Support Services team led by NCSA's Jay Alameda has helped the group get optimal performance out of their code, assisting with mapping processes to hardware and with I/O issues. The team's applied through XSEDE for 2 million CPU hours to be used to process the 1940 census records. ☐



PROJECT AT A GLANCE

VIRTUAL WORLD EXPLORATORIUM TEAM MEMBERS

Nosh Contractor (Northwestern)
Marshall Scott Poole (Illinois)
Jaideep Srivastava (Minnesota)
Dmitri Williams (USC)

FUNDING

National Science Foundation
Army Research Institute
Air Force Research Lab

FOR MORE INFORMATION

http://vwobservatory.org/

CENSUS PROJECT/ISDA TEAM MEMBERS

Frederico Bassetti
Devin Bonnie
Liana Diesendruck
Rob Kooper
Zhijin Li
Luigi Marini
Kenton McHenry
Michal Ondjrejcek
Victoria Winner

FUNDING

National Archives and Records Administration

FOR MORE INFORMATION

http://isda.ncsa.illinois.edu/

ACCESS ONLINE

www.ncsa.illinois.edu/News/Stories/bigdata

BLUE WATERS EARLY SCIENCE SYSTEM

SIX RESEARCH TEAMS HAVE BEGUN USING THE FIRST phase of the Blue Waters sustained-petascale supercomputer to study some of the most challenging problems in science and engineering, from supernovae to climate change to the molecular mechanism of HIV infection.

The Blue Waters Early Science System, which is made up of 48 Cray XE6 cabinets, represents about 15 percent of the total Blue Waters computational system and is currently the most powerful computing resource available through the National Science Foundation.

"This is an exciting and important milestone in the Blue Waters project," says Irene Qualters, program director of the NSF Office of Cyberinfrastructure. "It began as an idea, and now thanks to sustained collaborative efforts by the entire project team, the vendor, and the science teams, this computational tool is beginning to advance fundamental understanding in a wide range of scientific topics."

More than two dozen research teams have been awarded Petascale Computing Resource Allocations (PRAC) through a competitive NSF-led process. The PRAC awards enable these teams to work with NCSA to prepare their codes to take full advantage of Blue Waters and other extreme-scale computing systems. The teams submitted proposals outlining how they could use the Early Science System during the limited time it is available before being integrated into the full Blue Waters system; it was challenging to select just a few of the teams to achieve the first Blue Waters science results.

"All of these outstanding science and engineering teams are poised to do great, boundary-expanding work. The achievements of the first set of pioneers will soon be followed by those of their colleagues when the full system becomes available later this year," says NCSA Director Thom Dunning, principal investigator for the Blue Waters project.

The Early Science System research teams and their projects are:

- **Homayoun Karimabadi's team**, at the University of California-San Diego, is modeling high-temperature plasmas, including magnetic reconnection and flux transfer events to better understand the impact of the solar wind and solar flares on the Earth's atmosphere.
- **Brian O'Shea and his team**, Michigan State University, are simulating the formation and evolution of the Milky Way's most distant ancestors, a population of small galaxies formed shortly after the Big Bang. These simulations will be more accurate and full featured than any performed before. Previous simulations have modeled volumes of 1 megaparsec; the highest resolution subvolume of O'Shea's simulation will be more than 200 times this size.
- **Klaus Schulten and his team** at the University of Illinois at Urbana-Champaign are studying the protein capsid that encases the HIV-1 genome. The process through which this capsid disassembles, releasing its genetic material, is a critical step in HIV infection. Schulten's group will simulate a cylindrical capsid consisting of 12.5 million atoms.
- **Robert Sugar's team**, University of California, Santa Barbara, is conducting lattice quantum chromodynamics studies, which deal with sub-atomic physics. The first goal is to examine the charmonium spectrum (the bound state of charms and anti-charms); this requires challenging simulations with very small lattice spacing and very large lattice dimensions. The team's second effort deals with exotic mesons, which are prime territory in which to hunt for gluons. Sugar's team aims to use the Blue Waters Early Science System to confirm or refute that this state exists and can be identified experimentally.
- **Stanford Woosley's team**, University of California, Santa Cruz and Lawrence Berkeley Lab, is researching explosive burning in Type Ia supernovae, which are used as "standard candles" for surveying astronomically vast distances. Using the Early Science System, Woosley's team will be able to achieve unprecedented resolution at the finest level of their adaptive mesh refinement simulations.
- **Donald Wuebbles and his team**, University of Illinois at Urbana-Champaign, will simulate the end of both the 20th and 21st centuries at 0.25° global resolution. These high-resolution time slices will enable his team to explore changes in the frequency and intensity of extreme events, such as tropical cyclones and mid-continental thunderstorms that are not adequately resolved in global climate models at lower resolution. By using the Early Science System, Wuebbles' team hopes to contribute high-resolution results in time for the next assessment report of the Intergovernmental Panel on Climate Change.

For more information about Blue Waters and the science and engineering research it will support, see www.ncsa.illinois.edu/BlueWaters/.

ILLINOIS DATA CENTER RATED LEED GOLD

The University of Illinois' National Petascale Computing Facility has been certified LEED® Gold in the U.S. Green Building Council's Leadership in Energy and Environmental Design (LEED) rating system, which is the recognized standard for measuring sustainability in construction.

The building, which opened in summer 2010, is home to supercomputers and other high-performance systems operated by NCSA and used by scientists and engineers across the country.

The LEED system awards points in a half-dozen categories, including energy and water efficiency, indoor environmental quality, and innovation in design. NPCF's energy-saving features include:

A highly efficient power distribution system that is based on 480-volt power for the computational equipment.

A liquid cooling system for the computational and data storage equipment, which is two times more efficient than air cooling.

External cooling towers that let Mother Nature chill the water needed for cooling the building and the supercomputers a large part of the year. This is expected to cut the chilled water costs for the facility by about $1 million per year.

Low-impact landscaping with native prairie plants that thrive without frequent watering or mowing.

Using best practice construction methods to improve the air quality environment within the facility.

XSEDE12 CONFERENCE IN CHICAGO

The Extreme Science and Engineering Discovery Environment (XSEDE) is the most advanced, powerful, and robust collection of integrated advanced digital resources and services in the world. It is a single virtual system that scientists can use to interactively share computing resources, data, and expertise. XSEDE supports supercomputers and high-end visualization and data analysis resources across the country—as well as data collection and computational tools that are critical to the success of researchers—by integrating resources and services, making them easier to use.

The XSEDE annual conference will be held July 16-20 in Chicago. XSEDE12 will showcase the discoveries, innovations, and achievements of those who use, build, and support XSEDE resources, as well those not directly involved with XSEDE. XSEDE12 also will create a forum for discussion of current needs and future plans among researchers, students, XSEDE staff, and National Science Foundation representatives. The conference is currently accepting submissions for tutorials, panels, papers, and posters. For more information see: **www.xsede.org/xsede12**.

NCSA, ILLINOIS TO AID GENOMICS COMPETITION

The Archon Genomics X PRIZE presented by Medco is intended to inspire breakthroughs in genome sequencing that will lead to the creation of a "medical grade" genome that can be used to improve patient diagnosis and treatment. Beginning in January 2013, teams will compete to accurately sequence the genomes of 100 healthy centenarians within

30 days for less than $1,000 per genome. A $10 million prize will be awarded to either a single winner or divided among successful teams.

Determining who wins the $10 million prize is no easy feat. It will require a validation protocol crafted by a team of genomics experts, the development of sophisticated bioinformatics software, and the data-crunching power of the supercomputers at NCSA.

Step one was defining, for the first time, what it means to have a complete and accurate "whole human genome sequence." This validation protocol is not simply important in determining the competition's winner; it also creates a universal standard of quality, accuracy, and completeness for the sequencing industry. Among the experts who began work on this standard in 2010 was Victor Jongeneel, a senior research scientist at both NCSA and Illinois' Institute for Genomic Biology (IGB).

Once competitors complete their sequences, they will transfer their massive data sets to NCSA, which will ensure the confidentiality and proper handling of the data. Sophisticated bioinformatics software developed by EdgeBio will be run on NCSA's high-performance computing systems, which have the teraflops of computing power required to quickly analyze the data and determine the winner without controversy.

One hundred centenarians—known as the Medco 100 Over 100—are donating their DNA for this competition. Their genome sequences may provide valuable insights into health and longevity. When the contest is over, most of their data will be available to researchers via the National Institutes of Health's dbGaP data resource and NCSA will archive the raw data for the X-PRIZE Foundation.

Anyone 100 years of age or older can be nominated to participate in this research effort. To nominate a centenarian, go to: **http://genomics.xprize.org/medco-100-over-100.**

GLOBUS ONLINE SELECTED FOR BLUE WATERS DATA

NCSA selected Globus Online as the data movement solution for the Blue Waters supercomputer. A secure, reliable service for high-performance file transfer, Globus Online will be used by researchers across the country, both for remote file transfer into and out of the system and in some cases for movement of files within Blue Waters between its archival mass storage system and its high-speed parallel file system.

Globus Online is a cloud-hosted service for fast, reliable data movement that can be accessed through the Internet. It offers both a web interface and command line interface for file transfer, as well as a REST API for easy integration with existing systems. Globus Online automates the mundane (but error-prone and time-consuming) activity of managing file transfers between supercomputing facilities, cloud resources, campus clusters, lab servers, desktop and laptops. The process significantly reduces transfer time, with some users reporting movement of terabytes of data in hours.

Globus Online is an initiative by the Computation Institute at the University of Chicago and Argonne National Laboratory and is supported in part by funding from the Department of Energy, the National Science Foundation, and the National Institutes of Health. To get started or find out more, visit **www.globusonline.org**.

PENNINGTON RETURNS TO NCSA

After serving as a program director in the Office of Cyberinfrastructure at the National Science Foundation for the past few years, NCSA deputy director and chief Technology Officer Rob Pennington has returned to NCSA. He may be contacted at **rlpennin@illinois.edu.**

NCSA INTERN WINS AWARDS

Briana Chapman, a high school student who works with NCSA's Cybereducation Division, was named a runner up in the National Center for Women in Technology's Award for Aspirations in Computing contest. The NCWIT Award for Aspirations in Computing honors young women at the high-school level for their computing-related achievements and interests. Awardees are selected for their computing and IT aptitude, leadership ability, academic history, and plans for post-secondary education. In addition, she was named a winner in the Illinois Affiliate competition and will be honored in May at an event in Chicago.

NCSA'S HOEFLER HONORED

NCSA's Torsten Hoefler, who leads the application and system performance modeling and simulation efforts for the Blue Waters project, has been selected as the recipient of the 2012 SIAG/Supercomputing Junior Scientist Prize. The award honors distinguished contributions in the field of algorithms research and development for parallel scientific and engineering computing.

Hoefler's research revolves around performance-centric software development and deals with scalable networks, parallel programming techniques, and performance modeling.

He received the award at the 2012 SIAM Parallel Processing Conference, at which he gave a presentation about his research work title "Performance-oriented Parallel Programming: Integrating Hardware, Middleware and Applications."

For more information about Hoefler's research, see **www.unixer.de**.

AVL WORK HELPS EARN OSCAR NOMINATIONS

This early parameter study shows an exploration by AVL and Dan Glass of the Volker Bromm supernova data with added density and detail layers. After further visual development, final high resolution layers were provided to Double Negative Visual Effects, London for further processing.

The film didn't take home the coveted statue, but "The Tree of Life" was nominated for Best Picture for this year's Academy Awards. NCSA's Advanced Visualization Laboratory (AVL), led by Donna Cox, collaborated with director Terrence Malick, visual effects supervisor Dan Glass, and others to create two animated visualizations for the film that are based on scientific data.

In the film, a middle-aged man's contemplation of the pattern and meaning of his life is interwoven with moments from his childhood in a small Texas town. As part of the meaning-of-life sequences, the AVL team supplied the images conveying birth and death in the universe and the majesty of flying through the galaxy.

The collaboration included exploring the data and testing a variety of camera moves, visual looks, and approaches. The AVL team also developed new software to add detail and realism. They worked on a cluster of computers dedicated to the AVL with about 200 processors. When they needed more power, they used a 9,600-processor NCSA supercomputer called Abe, which is now retired.

The team also visualized a supercomputer simulation by Volker Bromm, an astronomy professor at the University of Texas at Austin. It shows how the very first stars appeared, illuminating the previously dark universe. Bromm simulated this event at the Texas Advanced Computing Center on their Ranger supercomputer over the course of 42 days. The calculation would have taken 114 years on a laptop.

"The Tree of Life," which starred Brad Pitt, also earned nominations for best director for Terrence Malick and best cinematography for Emmanuel Lubezki.

ADAPTIVE COMPUTING PARTNERS WITH NCSA

Adaptive Computing, manager of some of the world's largest supercomputing systems and an expert in high-performance computing (HPC) workload management, is the latest company to join NCSA's Private Sector Program, which puts the center's expertise to work on some of the toughest challenges faced by industry.

Providing Adaptive Computing's Moab intelligence engine on NCSA's iForge HPC system creates a premier computing environment for the center's industrial power users, allowing users to bench test code using an intelligence engine within a real-world environment. This partnership will also allow Adaptive Computing to advance its HPC workload management products and services to better serve the industrial market.

For more information on NCSA's industry partnership program, go to industry.ncsa.illinois.edu, or contact Merle Giles:
mgiles@ncsa.illinois.edu or 217-244-4629.

SPACE JUNK

A growing collection of debris is orbiting the earth, creating hazards that jeopardize both space exploration and the satellite network that powers modern communication systems. These cluttered orbits are the subject of the new film "Space Junk 3D," which features data-driven scientific visualizations created by NCSA's Advanced Visualization Laboratory (AVL).

Building on previous animations developed for NASA's James Webb Space Telescope project, AVL created two sequences based on scientific data from computer simulations. The first shows the evolution of filamentary structure in the early universe, using data from Princeton astrophysicists Renyue Cen and Jeremiah Ostriker. The second features a dramatic collision of galaxies created based on simulations conducted by Brant Robertson at the University of Arizona.

The AVL team designed these scenes using their ultra-high-resolution 3D visualization environment and two key pieces of software they have developed: Virtual Director for interactive scene design and Amore for rendering both volume and particle data. They also used one of NCSA's supercomputers to complete computationally demanding tasks by the production deadline.

"It's a real treat to be asked to participate in the making of 'Space Junk 3D'," says Bob Patterson, AVL senior research artist. "It's an opportunity to contribute cinematic scientific visualization to a giant screen science documentary to tell the larger story of natural phenomena in the cosmos."

The 38-minute film opened in January; it is shown in IMAX® and other giant screen theaters in 2D and 3D. Presented by Melrae Pictures, in association with Red Barn Productions, it's produced by Melissa Butts and Kimberly Rowe and distributed globally by K2 Communications.

PROJECT AT A GLANCE

AVL TEAM MEMBERS

Donna Cox
Robert Patterson
Stuart Levy
Matthew Hall
Jeff Carpenter

FOR MORE INFORMATION

http://avl.ncsa.uiuc.edu
www.spacejunk3d.com

IMAGE CREDIT

Data-driven visualization of colliding galaxies created at the National Center For Supercomputing Applications.

NCSA

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

1205 West Clark Street
Urbana, IL 61801

www.ncsa.illinois.edu